Effect of ripple repetition rate on discrimination of ripple glide direction and the detection of brief tones in spectro-temporal ripple noise

Vijaya Kumar Narne, Periannan Javahar Antony, Thomas Baer, and Brian C. J. Moore

Citation: The Journal of the Acoustical Society of America **145**, 2401 (2019); doi: 10.1121/1.5098770
View online: https://doi.org/10.1121/1.5098770
View Table of Contents: https://asa.scitation.org/toc/jas/145/4
Published by the Acoustical Society of America

Effect of ripple repetition rate on discrimination of ripple glide direction and the detection of brief tones in spectro-temporal ripple noise

Vijaya Kumar Narne[a)]
Department of Audiology, JSS Institute of Speech and Hearing, Mysore, India

Periannan Javahar Antony
Department of Audiology, All India Institute of Speech and Hearing, Mysore, India

Thomas Baer and Brian C. J. Moore
Department of Experimental Psychology, University of Cambridge, Cambridge, United Kingdom

(Received 18 July 2018; revised 25 March 2019; accepted 27 March 2019; published online 26 April 2019)

The effect of temporal repetition rate R on the discrimination and internal representation of stimuli with spectro-temporal ripples was examined. Experiment 1 measured the highest ripple density D at which upward- and downward-gliding ripples could be discriminated. Thresholds varied only slightly for R from 2 to 8 Hz, with a median threshold just above 5 ripples/oct. The threshold decreased (worsened) when R was increased to 16 and 32 Hz, suggesting that the limited temporal resolution of the auditory system plays a role for these higher values of R. Experiment 2 explored the internal representation of stimuli with static and downward-gliding spectral ripples by measuring the detection threshold for a brief tone presented at a peak or a valley in the stimulus spectrum. Thresholds were generally higher when the signal was at a peak than when it was at a valley. The peak-valley difference tended to decrease with increasing D, and the variation of thresholds with D was greater for low R than for high R. The results suggest that the discrimination of spectro-temporal ripples is limited mainly by frequency resolution for lower ripple rates (up to 4–8 Hz) but temporal resolution plays a major role for higher rates.
https://doi.org/10.1121/1.5098770

[VMR] Pages: 2401–2408

I. INTRODUCTION

Spectral ripple noise (SRn) is based on a "carrier" that is either white noise or multiple sinusoids that are usually uniformly spaced on a linear or logarithmic frequency scale. The spectrum of the carrier is modulated with a sinusoidal function, usually on a linear or logarithmic frequency scale (Supin *et al.*, 1998; Won *et al.*, 2007). The result is a broadband stimulus with regular amplitude fluctuations along the frequency axis (spectral modulations). For log-spaced ripples, the rate of spectral modulation, or ripple density, D, is usually specified in ripples per octave (RPO). The maximum value of D at which the ripples can be detected or at which two different ripple densities or ripple phases can be discriminated has often been used as a measure of spectral resolution for people with normal hearing (NH), cochlear hearing loss (CHL), and cochlear implants (CI) (Henry *et al.*, 2005; Supin *et al.*, 1994; Won *et al.*, 2007). In variants of these tests, the ripples have been made to glide upwards or downwards in frequency to produce a spectro-temporal ripple noise (STRn), with temporal repetition rate, R (in ripples/s). It has been assumed in many studies using STRn stimuli that performance depends on both spectral and temporal resolution (Archer-Boyd *et al.*, 2018; Chi *et al.*, 1999; Nechaev *et al.*, 2018). However, the relative importance of these two has not been systematically assessed. This paper examines the effect of R on performance of a task requiring discrimination of ripple glide direction and it also explores the internal representation of STRn stimuli as a function of D and R, by measuring the detection threshold for a brief tone presented at a peak or valley in the short-term spectrum of the STRn. The main goal was to assess the relative role of temporal and spectral resolution in determining performance in discriminating ripple glide direction.

We start by reviewing previous studies using SRn and STRn stimuli and discuss some problems with different variants of the tests. Azadpour and McKay (2012) showed that thresholds obtained using SRn stimuli with static ripples may not provide an accurate measure of spectral resolution for CI users, for whom even ripples with very low densities are hard to detect and discriminate. They reported that, in addition to spectral cues relating to the ripples, there were confounding cues, such as differences in loudness, spectral centroid (the weighted mean frequency), and changing energy at the spectral edges, that led to better thresholds than would be obtained if only cues from the spectral ripples were used. To avoid the use of such confounding cues Aronoff and Landsberger (2013) modified the SRn stimulus to produce an STRn in which the phase of the spectral ripples changed over time, effectively making the ripples glide upwards or downwards in frequency. The task, called the

[a)]Also at: Institute of Clinical Research, South Denmark University, Odense, Denmark. Electronic mail: vijaynarne@gmail.com

STRt, was to discriminate a test stimulus with adjustable D from a reference stimulus with a high value of D, such that the spectral ripples were completely unresolved. The value of R was 5 Hz. As a result of the temporal changes, the overall loudness evoked in each local frequency region, the energy at the spectral edges, and the spectral centroid became unreliable cues for discriminating the target and reference stimuli. However, Narne *et al.* (2016) showed that the results for the STRt could be affected by a confounding cue, namely, amplitude fluctuations with rate R at the outputs of auditory filters tuned above and below the passband of the stimuli, which occur even when R is sufficiently high that such fluctuations are negligible at the outputs of auditory filters tuned within the passband of the STRt.

A variant of the STRt is the $STRt_{dir}$ described by Narne *et al.* (2018), which estimates the highest value of D at which an upward-gliding ripple can be discriminated from a downward-gliding ripple. This test was intended to avoid both the confounding cues described by Azadpour and McKay (2012) for stimuli with static spectral ripples and the confounding cues described by Narne *et al.* (2016) for stimuli with gliding spectral ripples. The cues become less usable in the latter case because the upward- and downward-gliding ripples exhibit identical long-term modulation spectra in frequency regions near the lower and upper edges of the passband of the stimuli.

As noted earlier, performance of the STRt and $STRt_{dir}$, and related tests that require the discrimination of glide direction (Archer-Boyd *et al.*, 2018), may depend on both spectral and temporal resolution, depending on the value of R. For low values of R, spectral resolution is probably the dominant factor. However, as R increases, temporal resolution plays an increasing role, and for very high R temporal resolution is likely to be the main factor limiting performance. This is analogous to the temporal modulation transfer function (TMTF) for detection of amplitude modulation (AM) (Viemeister, 1979). For AM rates below 80–100 Hz, performance depends mainly on the intensity resolution of the auditory system and the threshold is almost independent of AM rate, but for higher AM rates, temporal resolution plays a greater role, and the threshold for detecting AM increases. For very high rates (>1000 Hz), the AM is essentially undetectable when the carrier frequency is high and the spectral sidebands are not resolved, although performance can improve with increasing AM rate when the sidebands are resolved (Kohlrausch *et al.*, 2000; Moore and Glasberg, 2001). In the case of the STRt, we are not aware of any studies assessing the effect of R. Following Aronoff and Landsberger (2013), several studies have used the STRt with $R = 5$ (Kirby *et al.*, 2015; Narne *et al.*, 2016; Vickers *et al.*, 2016; Zhou, 2017). However, no clear rationale was given for the choice of R, and it is not clear whether temporal resolution influences the detectability of spectral ripples for this value of R.

There have been some studies exploring the effect of R on spectral modulation sensitivity assessed with tests other than the STRt or $STRt_{dir}$, using participants with NH (Bernstein *et al.*, 2013; Chi *et al.*, 1999; Mehraei *et al.*, 2014) and CHL (Bernstein *et al.*, 2013; Mehraei *et al.*, 2014). These studies assessed the smallest spectral modulation depth (in dB) required to discriminate a STRn with ripple density D and ripple rate R from an unmodulated stimulus. Thresholds tended to increase when R was increased above about 4 Hz, suggesting an influence of temporal resolution for R above 4 Hz. Nechaev *et al.* (2018) estimated the highest spectral ripple velocity (in oct/s, defined as R/D) at which an STRn stimulus could be discriminated from a noise without spectral ripples, for various values of D of the STRn. They found that the ripple velocity at threshold decreased from 389 oct/s for $D = 1$ ripple/oct to 11.3 oct/s for $D = 7$ ripples/oct. They hypothesized that performance was limited both by the filtering that occurs in the auditory periphery and by a central temporal integrator (Moore *et al.*, 1988; Plack and Moore, 1990), the relative importance of the two depending on ripple density.

The main goal of the present paper is to assess the relative importance of spectral and temporal resolution in the perception of STRt stimuli for different values of R. Experiment 1 assessed the highest value of D at which an upward-gliding ripple could be discriminated from a downward-gliding ripple, with values of R from 2 to 32 Hz. Experiment 2 explored the internal representation of STRn stimuli by measuring the detection threshold for a brief tone burst placed at peak or a valley of a STRn, using values of D from 1 to 4 and values of R from 0 (static ripples) to 32 Hz. The results suggest that spectral resolution plays a dominant role for lower values of R, but that temporal resolution plays a dominant role for higher values of R.

II. EXPERIMENT 1: DISCRIMINATION OF UPWARD- AND DOWNWARD- GLIDING RIPPLES AS A FUNCTION OF *R*

A. Method

1. Participants

Ten normal-hearing participants (3 females and 7 males) with a mean age of 23 years (range: 18–32 years, standard deviation, SD = 5.2 years) were tested. Pure-tone thresholds and speech identification scores were measured using a calibrated diagnostic audiometer (AD-226, Interacoustics, Middelfart, Denmark) and Sennheiser HDA200 headphones (Old Lyme, CT, USA). All participants had pure-tone thresholds ≤15 dB hearing level (HL) for octave frequencies between 0.25 and 8 kHz. All had identification scores greater than 90% for words in quiet, presented at 40 dB above the pure-tone average (PTA) threshold across 0.5, 1, 2, and 4 kHz. All participants had normal tympanograms, with ipsilateral acoustic reflexes at normal levels in both ears, indicating normal middle ear function. None of the participants had a history of neurological or otological disorder.

All testing procedures were approved by the JSS Institute of Speech and Hearing review board, and written informed consent was obtained from all participants.

2. Stimuli

The STRn was generated exactly as described by Narne *et al.* (2016). The carrier was composed of 201 equal-amplitude

sinusoidal frequency components, spaced every 0.03 octave from 100 to 6400 Hz. The amplitude of each of the sinusoidal component was modulated as a function of D and R. The STRn stimulus as a function of time, STRn(t), for downward-gliding ripples is given by Eqs. (1) and (2),

$$STRt(t) = \sum_{i=1}^{i=201} \frac{P(i)}{201} \times 10^{\frac{d\left(|\sin(\pi Rt + 0.03i\pi D + \theta_2)| - 1\right)}{20}}, \quad (1)$$

where

$$P(i) = \sin(2\pi f_c(i) + \theta_1(i)). \quad (2)$$

$P(i)$ is the ith carrier component with frequency $f_c(i) = 100 \times 2^{0.03(I-1)}$ and $\theta_1(i)$ is the starting phase of that component, which was chosen randomly for each component and each stimulus over the range 0–2π radians. The phase of the spectral ripple at the onset of the stimulus, θ_2, was chosen randomly for each stimulus over the range 0–π radians. The ripple depth d is the peak-to-trough ratio in dB, which was kept constant at 20 dB. The spectral ripples were roughly sinusoidal in form on a linear amplitude scale and a logarithmic frequency axis. To generate the upward-gliding ripples, R was multiplied by -1. Values of R were 2, 4, 8, 16, and 32 Hz. In addition, a condition was included using static ripples with randomized ripple starting phase, in which the task was to discriminate a test SRn with variable D from a reference SRn with a nominal value of $D = 20$ RPO. For convenience, this condition is described as $R = 0$. Because there were 33.33 carrier components per octave, the spectral ripples were not adequately sampled when the nominal value of D was 20. The resulting spectrum contained a mixture of two ripple rates: 13.33 and 6.67 RPO. However, excitation patterns for this stimulus, calculated using the method described by Glasberg and Moore (1990), were smooth, and were unaffected by the value of θ_2, indicating that the ripples were completely unresolved. During the adaptive procedure described below, the value of D for the test SRn never exceeded 9 RPO for $R = 0$, and never exceeded 8 for $R = 2$–32 Hz, so the ripples were always adequately sampled by the carrier components. Hence no data were rejected because of inadequate sampling of the spectral ripples. The overall level of the stimuli was 75 dB sound pressure level (SPL).

The stimuli were generated, and responses were collected via a personal computer and a 32-bit sound card (Realtek, Hsinchu, Taiwan) using a sampling rate of 22 050 Hz. Stimuli were routed through an audiometer (Interacoustics AD-226) to one earpiece of the headphones (Sennheiser HDA200). Five participants were tested using the right ear and five using the left ear.

3. Procedure

All participants were seated in a double-walled air-conditioned sound-treated room. Thresholds were estimated using a three-alternative forced-choice (3-AFC) task. The participant was instructed to indicate the interval that differed from the other two intervals. For $R = 0$ (static ripples), three successive stimulus bursts were presented in each trial, each with a duration of 750 ms, including 100-ms raised-cosine rise/fall ramps. The silent interval between bursts was 500 ms. Two bursts were "standards" with a nominal $D = 20$ RPO and the other one was a "target" with variable D. For $R \neq 0$, there were three pairs of stimulus bursts in each trial, giving six successive bursts in total, each with a duration of 750 ms, including 100-ms raised-cosine rise/fall ramps. Within each pair the silent interval was 150 ms, and the silent interval between pairs was 500 ms. Of the three pairs, two were standards and one was the target. The two standard pairs both containing two downward-gliding STRn bursts. The target pair contained a downward-gliding STRn burst followed by an upward-gliding burst. For all values of R, the position of the target was randomized across trials. Feedback indicating the correct interval was given on each trial, after the participant had responded.

The value of D was initially set to 1 RPO. D was increased following two correct responses and decreased following one incorrect response, so as to track the 70.7% correct point on the psychometric function (Levitt, 1971). The initial step size was 0.5 RPO and after two reversals it was decreased to 0.2 RPO. Eight reversals were obtained, and the threshold was taken as the arithmetic mean of the values of D at the last six. The value of D never reached 0. All participants were given two practice runs before commencing the experiment, using $R = 4$ and $R = 16$ Hz. We have found previously that this is enough to give stable threshold estimates. Only a single threshold estimate was obtained for each condition and each participant, due to limited availability of the participants. Although the variability of the threshold estimate for a single participant might be fairly high, we focus on the median results across the ten participants. The order of stimulus conditions was randomized across participants.

B. Results

Figure 1 shows the thresholds obtained using the $STRt_{dir}$ as a function of R. The results for the condition with static ripples are plotted at $R = 0$. A one-way repeated-measures analysis of variance (ANOVA) on the data for $R = 2$–32 Hz showed a significant effect of R ($F_{(4,36)} = 181.72$, $p < 0.001$, $\xi^2 = 0.85$). Bonferroni-corrected pairwise comparisons showed that thresholds were significantly lower (worse) for $R = 16$ and 32 Hz than for all other values of R. The value of D at threshold was just above 5 RPO for $R \leq 8$ Hz, consistent with the results of Narne *et al.* (2018) for $R = 5$ Hz. The median threshold across participants for the condition with static ripples was similar to that reported by Narne *et al.* (2016) and was also similar to the median threshold across participants for the conditions with $R = 2$–8 Hz.

C. Discussion

The similarity of thresholds for R from 2 to 8 Hz and the worsening in performance for higher values of R resembles the pattern of results observed for the discrimination of amplitude-modulation phase (Dau, 1996; Thompson and Dau, 2008), which is simulated in the model of Dau *et al.* (1997a,b) by assuming that modulation phase information is discarded for modulation filters tuned above 8 Hz.

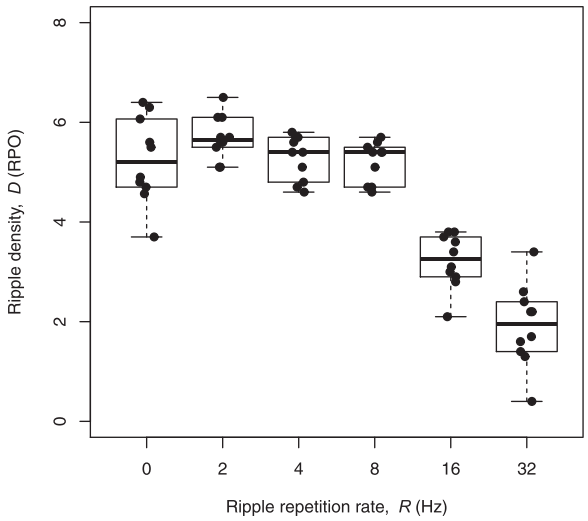

FIG. 1. Box plots showing the value of D at threshold for discriminating upward-gliding from downward-gliding spectral ripples as a function of ripple repetition rate, R (2–32 Hz), as obtained in experiment 1. The individual thresholds are shown by circles. The thick horizontal lines show medians, the lower and upper edges of the boxes show the first and third quartiles, and the whiskers indicate the 10th and 90th percentiles. The data plotted at $R = 0$ are thresholds for a task that required discrimination of a static ripple noise with variable density from a reference ripple noise with $D = 20$ (for which the ripples were assumed to be completely unresolved).

The worsening in performance for $R = 16$ and 32 Hz suggests that performance of the $STRt_{dir}$ is partly limited by temporal resolution for these rates. The worsening may reflect at least two factors. First, it is possible that frequency selectivity takes some time to develop and that it is reduced for stimuli that are very brief or whose spectra fluctuate rapidly. This might explain, for example, why the threshold for detecting a brief tone pulse presented in a broadband noise masker is higher when the tone pulse is presented just after the masker onset than when it is presented later on in the masker (Zwicker, 1965). There is controversy about whether the sharpness of the auditory filter takes time to build up. Some studies appear to show a build-up of frequency selectivity with increasing presentation time of a masker over a few tens of milliseconds (Bacon and Moore, 1986a,b, 1987; Bacon and Viemeister, 1985a) while others show no change of the auditory filter over time (Moore *et al.*, 1987). It has been argued that peripheral filtering processes do not develop over time and that the temporal effects, when observed, depend on more central processes related to auditory change detection (Bacon and Moore, 1987; Zwicker and Fastl, 1972).

Another factor that may contribute to the worsening in performance for $R = 16$ and 32 Hz is the existence of temporal integration processes occurring more centrally. It has been proposed that the limited temporal resolution of the auditory system can be modeled using a sliding temporal integrator or window with an equivalent rectangular duration of 8–10 ms (Moore *et al.*, 1988; Oxenham and Moore, 1994; Penner *et al.*, 1972). For high values of R, the internal representation of the spectro-temporal pattern of the stimulus would become somewhat blurred by the temporal smearing produced by the sliding temporal integrator, effectively reducing the peak-to-valley ratio of the internal representation. This idea was examined further in experiment 2, which explored the internal representation of STRn stimuli by measuring the detection threshold for a brief probe tone presented at either a peak or a valley in the spectro-temporal stimulus.

III. EXPERIMENT 2: DETECTION OF BRIEF TONE PULSES IN STRN

A. Method

1. Participants

Six NH participants (2 females and 4 males) with a mean age of 23 years [range: 18–32 years, standard deviation (SD) = 8.2 years] were tested. None took part in experiment 1, but all met the same criteria as for experiment 1.

2. Stimuli

The signal was a 3.2-kHz sinusoid with 5-ms raised-cosine on/off ramps and no steady state; the duration at the −6 dB points on the envelope was 5 ms. This duration was chosen to be short relative to the repetition period ($1/R$) of the STRn masker, which was 31 ms for the highest value of R used (32 Hz), while avoiding audible spectral splatter (Bacon and Viemeister, 1985b).The signal frequency was chosen so that when the starting phase of the ripple in the STRn [θ_2 in Eq. (1)] was set to 0 the signal coincided with a valley in the spectrum (left panel of Fig. 2), while when the

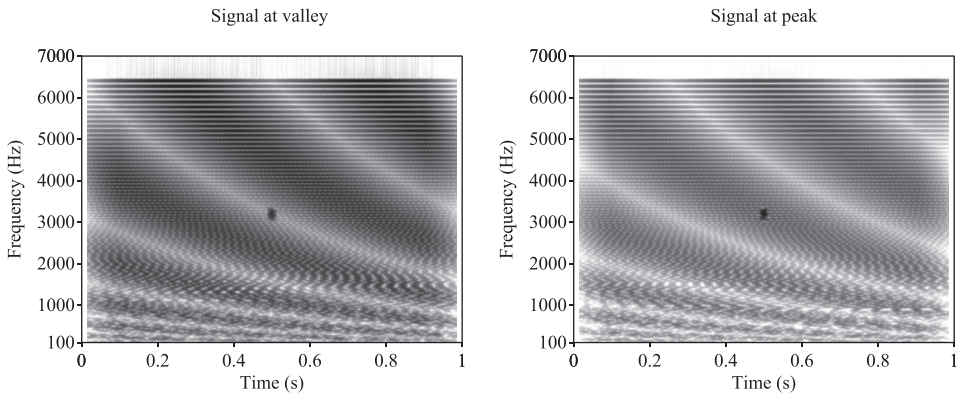

FIG. 2. Spectrograms of stimuli for experiment 2 for an example case with $D = 1$ RPO and $R = 2$ Hz. Darker areas indicate regions with higher energy. The brief 3.2-kHz signal was presented either at a spectral valley (left) or a spectral peak (right).

starting phase was $\pi/2$ the signal coincided with a peak in the spectrum (right panel of Fig. 2). The relatively high signal frequency also helped to ensure that spectral splatter was not audible (Bacon and Viemeister, 1985b). The STRn masker was presented in three bursts, separated by 500 ms, each with a duration of 1000 ms, including 100-ms raised-cosine ramps. The STRn was generated as described for experiment 1, but using only downward-gliding ripples. The level of the STRn was kept constant at 60 dB SPL and the level of the signal was varied to estimate the threshold. The level of the STRn was chosen to be lower here than for experiment 1, to avoid very high levels of the brief signal during the adaptive procedure. The signal was added to one of the three masker bursts, starting 495 ms after the onset of the masker. Thresholds for the signal at peak and signal at valley were measured for all combinations of values of D (1, 2, 3, and 4 RPO) and R (0, 2, 4, 8, 16, and 32 Hz) in a randomized block design: the six R values were tested in random order that was different for each participant, and for each value of R the four D values were tested in a random order. For each participant, all D and R conditions were completed for one signal location (signal at peak or valley) before testing with the other signal location. The order of testing of signal location was randomized across participants.

3. Procedure

Thresholds were measured using a three-interval 3-AFC procedure. The masker was presented in all three intervals while the signal was presented in one randomly selected interval. The task was to identify which interval contained the signal. Feedback as to the correct interval was provided after each response, via the screen of the PC. Each run started with the signal level set to 70 dB SPL. The signal level was adjusted adaptively for subsequent presentations using a three-down, one-up rule tracking the 79.4% correct point on the psychometric function (Levitt, 1971). The step size was 4 dB until the first reversal occurred and 2 dB thereafter. Eight reversals were obtained using the 2-dB step size and the arithmetic mean of the levels at the last six reversals was taken as the threshold. Participants were given practice using $R = 0$, 4, and 16 Hz with a few randomly selected values of D. Following this, a single threshold estimate was obtained for each condition for each participant, due to limited availability of the participants.

Inspection of the results indicated that for some conditions one or two participants gave thresholds that were markedly higher than for the other participants. The standard deviation of the signal level at the last six reversals tended to be large (3 dB or more) in cases where the thresholds were unexpectedly high, perhaps indicating lapses of attention or "forgetting what to listen for." To avoid an undue influence of these outliers, the results presented below are based on the median threshold across participants for each condition, rather than the mean.

B. Results

Figure 3 shows thresholds for the signal at a peak (top) and valley (bottom) in the short-term spectrum, plotted as a

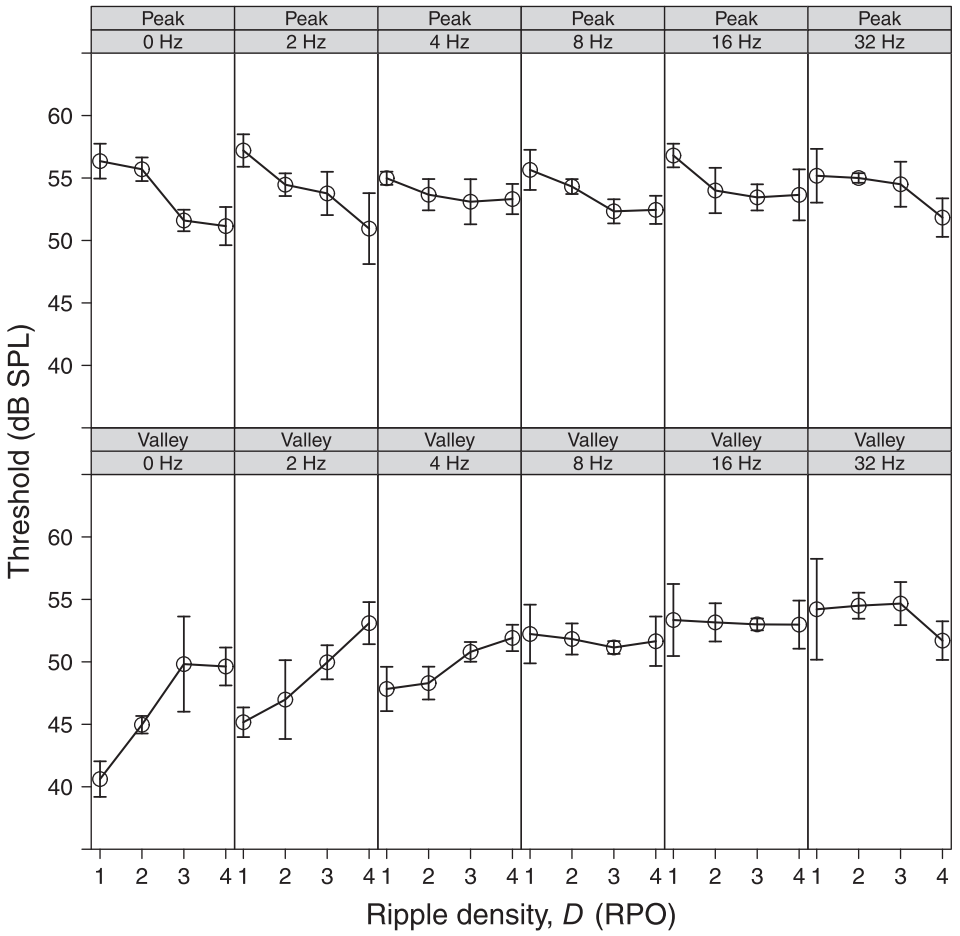

FIG. 3. The open circles show the median thresholds obtained in experiment 2 for detection of a brief 3.2-kHz signal presented at a peak (top) or a valley (bottom) in the spectrum, plotted as a function of ripple density, D. Each panel represents one ripple repetition rate R, as indicated at the top of the panel. Error bars show 95% confidence intervals.

function of ripple density, D. Each panel shows thresholds for one value of R. As expected, thresholds were generally higher when the signal was at a peak than when it was at a valley, although this was not always the case for $D = 3$ and 4. The variation of thresholds with D tended to be greater for low R than for high R. For the signal at a peak, the thresholds for $D = 1$ and 2 did not change markedly with R, while for the signal at a valley the thresholds increased with increasing R for all D.

Figure 4 shows the threshold differences between signal at peak and signal at valley of the masker as a function of D for each R. The threshold differences generally decreased as D increased, but the decrease was smaller for higher values of R. A repeated-measures ANOVA was performed on the threshold differences with factors D and R. There were significant main effects of D ($F_{(3,72)} = 36.6$, $p < 0.001$, $\xi^2 = 0.41$) and R ($F_{(5,72)} = 20.1$, $p < 0.001$, $\xi^2 = 0.1$), and a significant interaction ($F_{(15,72)} = 6.1$, $p < 0.001$, $\xi^2 = 0.23$). To explore the nature of the interaction, a separate one-way repeated-measures ANOVA with factor D was performed for each R. To allow for multiple comparisons, the Sidak correction was used, but with one-tailed significance levels, since we were testing the hypothesis that the threshold differences would decrease with increasing D. This gave a corrected significance level of 0.017. The effect of D was significant for $R = 0$ ($F_{(3,20)} = 13.4$, $p < 0.001$, $\xi^2 = 0.77$) and $R = 2$ Hz ($F_{(3,20)} = 15.0$, $p < 0.001$, $\xi^2 = 0.79$), marginally significant for $R = 4$ ($F_{(3,20)} = 8.13$, $p = 0.02$, $\xi^2 = 0.67$), and not significant for $R = 8$–32 Hz ($F_{(3,20)} = 2.37$ to 0.4, $p = 0.1$ to 0. 91).

C. Discussion

For $R = 16$ and 32 Hz, the threshold differences hardly varied with D, suggesting that the internal peak-to-valley ratio did not depend on spectral resolution and was presumably limited by temporal resolution. This suggests that performance of the $STRt_{dir}$ was also limited by temporal resolution for these high values of R. However, for $R = 0$, 2, and 4 Hz, the threshold differences decreased with increasing D, suggesting an important influence of spectral resolution. The pattern of results can be interpreted in the following way. Because of the operation of the sliding temporal integrator described earlier, the auditory system cannot select only the portion of the masker that is coincident in time with the signal. Rather, the "effective" masker results from a smearing of the masker over time. As noted earlier, the equivalent rectangular duration of the temporal integrator is usually assumed to be about 8–10 ms (Moore *et al.*, 1988; Oxenham and Moore, 1994; Penner *et al.*, 1972), although temporal smearing can occur over longer durations than this because of the assumed sloping skirts of the temporal integrator. As the value of R increases, the effective masker when the signal is at a valley in the masker spectrum includes progressively more energy from adjacent temporal segments of the masker, whose short-term level is higher. This leads to increases in threshold with increasing R. Conversely, when the signal is at a peak, the inclusion of portions of the masker adjacent to the peak has little effect on performance, because those adjacent portions have lower energy than at the peak.

IV. GENERAL DISCUSSION

The effect of D on the threshold differences shown in Fig. 4 clearly decreased as R was increased from 2 to 8 Hz, suggesting that the contrast in the internal representation of the masker decreased with increasing R over this range, while the thresholds for the discrimination of ripple glide direction measured in experiment 1 hardly changed when R was increased from 2 to 8 Hz. This discrepancy might reflect several factors. In experiment 2, signal detection depended on selective listening to a small time-frequency segment of the signal, while in experiment 1 information about the direction of the glide was available over a wide range of times and center frequencies. The ability to integrate information over time and frequency may partially offset the deleterious effect of the temporal smearing produced by the sliding temporal integrator. In addition, in experiment 1, the effect of any reduction in contrast of the internal representation of the spectrum with increasing R might have been partially offset by the greater number of ripple cycles occurring in the fixed duration. Finally, the reduced peak-valley threshold difference as R was increased from 2 to 8 Hz in experiment 2 might partly reflect the limited ability of the participant to select to the optimal time/place to listen.

Another possibility is that the discrimination of glide direction does not depend solely on spectral resolution, but also reflects the use of temporal fine structure (TFS) information,

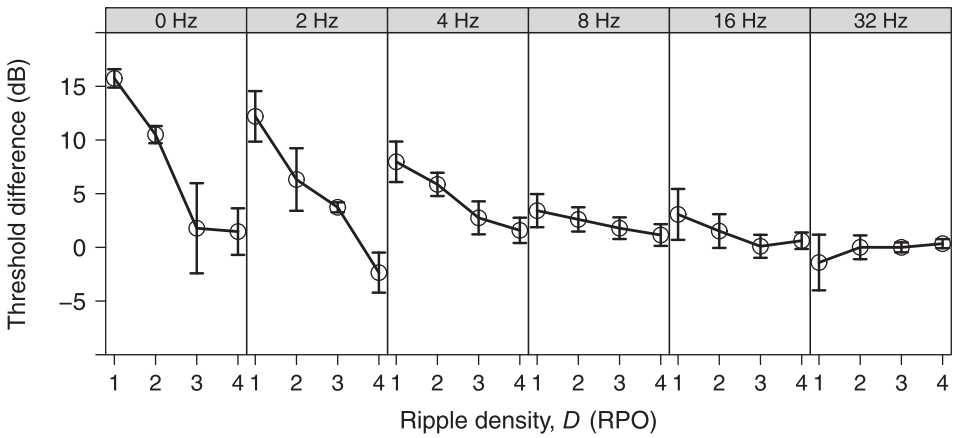

FIG. 4. The open circles show the median peak-valley threshold differences obtained in experiment 2. Each panel represents one ripple repetition rate R, as indicated at the top of the panel. Error bars show 95% confidence intervals.

specifically changes in TFS over time. There is good evidence that TFS information can contribute to the discrimination of stimuli with spectral ripples that are uniformly spaced on a linear frequency scale (Yost *et al.*, 1996) and there is some evidence suggesting that TFS information plays a role in the discrimination of stimuli similar to those used in experiment 1 (Bernstein *et al.*, 2013).

V. SUMMARY AND CONCLUSIONS

The main findings from the two experiments are summarized below.

(1) Experiment 1 measured the highest ripple density D at which upward- and downward-gliding spectral ripples could be discriminated. Thresholds varied only slightly for temporal repetition rates, R, from 2 to 8 Hz, with a median threshold just above 5 RPO, suggesting that thresholds over this range of R are determined mainly by spectral resolution. The value of D at threshold decreased (worsened) when R was increased to 16 and 32 Hz, suggesting that the limited temporal resolution of the auditory system plays a role at these higher values of R.

(2) Experiment 2 explored the internal representation of stimuli with static and downward-gliding spectral ripples by measuring the detection threshold for a brief tone presented at a peak or a valley in the stimulus. Thresholds were generally higher when the signal was at a peak than when it was at a valley. The peak-valley threshold difference was close to zero for all values of D for $R = 16$ and 32 Hz, consistent with the idea that ripple-direction discrimination for these high values of R was largely limited by temporal rather than spectral resolution.

ACKNOWLEDGMENTS

V.K.N. extends his appreciation for support to Professor N. P. Nataraja, Director, JSS Institute of Speech and Hearing. The authors would also like to acknowledge the participants for their co-operation. We also thank three reviewers for helpful comments on earlier versions of this paper.

Archer-Boyd, A. W., Southwell, R. V., Deeks, J. M., Turner, R. E., and Carlyon, R. P. (**2018**). "Development and validation of a spectro-temporal processing test for cochlear-implant listeners," J. Acoust. Soc. Am. **144**, 2983–2997.

Aronoff, J. M., and Landsberger, D. M. (**2013**). "The development of a modified spectral ripple test," J. Acoust. Soc. Am. **134**, EL217–EL222.

Azadpour, M., and McKay, C. M. (**2012**). "A psychophysical method for measuring spatial resolution in cochlear implants," J. Assoc. Res. Otolaryngol. **13**, 145–157.

Bacon, S. P., and Moore, B. C. J. (**1986a**). "Temporal effects in masking and their influence on psychophysical tuning curves," J. Acoust. Soc. Am. **80**, 1638–1645.

Bacon, S. P., and Moore, B. C. J. (**1986b**). "Temporal effects in simultaneous pure-tone masking: Effects of signal frequency, masker/signal frequency ratio, and masker level," Hear. Res. **23**, 257–266.

Bacon, S. P., and Moore, B. C. J. (**1987**). "Transient masking and the temporal course of simultaneous tone-on-tone masking," J. Acoust. Soc. Am. **81**, 1073–1077.

Bacon, S. P., and Viemeister, N. F. (**1985a**). "Simultaneous masking by gated and continuous sinusoidal maskers," J. Acoust. Soc. Am. **78**, 1220–1230.

Bacon, S. P., and Viemeister, N. F. (**1985b**). "The temporal course of simultaneous tone-on-tone masking," J. Acoust. Soc. Am. **78**, 1231–1235.

Bernstein, J. G. W., Mehraei, G., Shamma, S., Gallun, F. J., Theodoroff, S. M., and Leek, M. R. (**2013**). "Spectrotemporal modulation sensitivity as a predictor of speech intelligibility for hearing-impaired listeners," J. Am. Acad. Audiol. **24**, 293–306.

Chi, T., Gao, Y., Guyton, M. C., Ru, P., and Shamma, S. (**1999**). "Spectro-temporal modulation transfer functions and speech intelligibility," J. Acoust. Soc. Am. **106**, 2719–2732.

Dau, T. (**1996**). "Modeling auditory processing of amplitude modulation," Ph.D. thesis, University of Oldenburg, Germany.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997a**). "Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers," J. Acoust. Soc. Am. **102**, 2892–2905.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997b**). "Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration," J. Acoust. Soc. Am. **102**, 2906–2919.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Henry, B. A., Turner, C. W., and Behrens, A. (**2005**). "Spectral peak resolution and speech recognition in quiet: Normal hearing, hearing impaired, and cochlear implant listeners," J. Acoust. Soc. Am. **118**, 1111–1121.

Kirby, B. J., Browning, J. M., Brennan, M. A., Spratford, M., and McCreery, R. W. (**2015**). "Spectro-temporal modulation detection in children," J. Acoust. Soc. Am. **138**, EL465–EL468.

Kohlrausch, A., Fassel, R., and Dau, T. (**2000**). "The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers," J. Acoust. Soc. Am. **108**, 723–734.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Mehraei, G., Gallun, F. J., Leek, M. R., and Bernstein, J. G. W. (**2014**). "Spectrotemporal modulation sensitivity for hearing-impaired listeners: Dependence on carrier center frequency and the relationship to speech intelligibility," J. Acoust. Soc. Am. **136**, 301–316.

Moore, B. C. J., and Glasberg, B. R. (**2001**). "Temporal modulation transfer functions obtained using sinusoidal carriers with normally hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **110**, 1067–1073.

Moore, B. C. J., Glasberg, B. R., Plack, C. J., and Biswas, A. K. (**1988**). "The shape of the ear's temporal window," J. Acoust. Soc. Am. **83**, 1102–1116.

Moore, B. C. J., Poon, P. W., Bacon, S. P., and Glasberg, B. R. (**1987**). "The temporal course of masking and the auditory filter shape," J. Acoust. Soc. Am. **81**, 1873–1880.

Narne, V. K., Prabhu, P. P., Van Dun, B., and Moore, B. C. J. (**2018**). "Ripple glide direction discrimination and its relationship to frequency selectivity estimated using notched noise," Acta Acust. Acust. **104**, 1063–1074.

Narne, V. K., Sharma, M., Dun, B. V., Bansal, S., Prabhu, L., and Moore, B. C. J. (**2016**). "Effects of spectral smearing on performance of the spectral ripple and spectro-temporal ripple tests," J. Acoust. Soc. Am. **140**, 4298–4306.

Nechaev, D. I., Milekhina, O. N., and Supin, A. Y. (**2018**). "Hearing sensitivity to gliding rippled spectrum patterns," J. Acoust. Soc. Am. **143**, 2387–2393.

Oxenham, A. J., and Moore, B. C. J. (**1994**). "Modeling the additivity of nonsimultaneous masking," Hear. Res. **80**, 105–118.

Penner, M. J., Robinson, C. E., and Green, D. M. (**1972**). "The critical masking interval," J. Acoust. Soc. Am. **52**, 1661–1668.

Plack, C. J., and Moore, B. C. (**1990**). "Temporal window shape as a function of frequency and level," J. Acoust. Soc. Am. **87**, 2178–2187.

Supin, A. Y., Popov, V. V., Milekhina, O. N., and Tarakanov, M. B. (**1994**). "Frequency resolving power measured by rippled noise," Hear. Res. **78**, 31–40.

Supin, A. Y., Popov, V. V., Milekhina, O. N., and Tarakanov, M. B. (**1998**). "Ripple density resolution for various rippled-noise patterns," J. Acoust. Soc. Am. **103**, 2042–2050.

Thompson, E. R., and Dau, T. (**2008**). "Binaural processing of modulated interaural level differences," J. Acoust. Soc. Am. **123**, 1017–1029.

Vickers, D., Degun, A., Canas, A., Stainsby, T., and Vanpoucke, F. (**2016**). "Deactivating cochlear implant electrodes based on pitch information for users of the ACE strategy," Adv. Exp. Med. Biol. **894**, 115–123.

Viemeister, N. F. (**1979**). "Temporal modulation transfer functions based upon modulation thresholds," J. Acoust. Soc. Am. **66**, 1364–1380.
Won, J. H., Drennan, W. R., and Rubinstein, J. T. (**2007**). "Spectral-ripple resolution correlates with speech reception in noise in cochlear implant users," J. Assoc. Res. Otolaryngol. **8**, 384–392.
Yost, W. A., Patterson, R., and Sheft, S. (**1996**). "A time domain description for the pitch strength of iterated rippled noise," J. Acoust. Soc. Am. **99**, 1066–1078.
Zhou, N. (**2017**). "Deactivating stimulation sites based on low-rate thresholds improves spectral ripple and speech reception thresholds in cochlear implant users," J. Acoust. Soc. Am. **141**, EL243–EL248.
Zwicker, E. (**1965**). "Temporal effects in simultaneous masking by white-noise bursts," J. Acoust. Soc. Am. **37**, 653–663.
Zwicker, E., and Fastl, H. (**1972**). "On the development of the critical band," J. Acoust. Soc. Am. **52**, 699–702.